


RESEARCH ARTICLE



Characterization of the binding features between SARS-CoV-2 5'-proximal transcripts of genomic RNA and nucleocapsid proteins

Shih-Cheng Chen ^{a,b}, Cui-Ting Xu^b, Chuan-Fu Chang^b, Chia-Shin Yang^{c*}, Pin-Han Lin^d, Wei-Min Liu ^{d*}, Yeh Chen ^{c*}, and Chien-Hung Yu ^{b,e}

^aNational Institute of Cancer Research, National Health Research Institutes, Tainan City, Taiwan; ^bDepartment of Biochemistry and Molecular Biology, College of Medicine, National Cheng Kung University, Tainan City, Taiwan; ^cInstitute of Translational Medicine and New Drug Development, College of Medicine, China Medical University, Taichung City, Taiwan; ^dDepartment of Chemistry, Fu Jen Catholic University, New Taipei City, Taiwan; ^eInstitute of Basic Medical Sciences, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

ABSTRACT

Packaging signals (PSs) of coronaviruses (CoVs) are specific RNA elements recognized by nucleocapsid (N) proteins that direct the selective packaging of genomic RNAs (gRNAs). These signals have been identified in the coding regions of the nonstructural protein 15 (Nsp 15) in CoVs classified under *Embecovirus*, a subgenus of betacoronaviruses (beta-CoVs). The PSs in other alpha- and beta-CoVs have been proposed to reside in the 5'-proximal regions of gRNAs, supported by comprehensive phylogenetic evidence. However, experimental data remain limited. In this study, we investigated the interactions between Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) 5'-proximal gRNA transcripts and N proteins using electrophoretic mobility shift assays (EMSAs). Our findings revealed that the *in vitro* synthesized 5'-proximal gRNA transcripts of CoVs can shift from a major conformation to alternative conformations. We also observed that the conformer comprising multiple stem-loops (SLs) is preferentially bound by N proteins. Deletions of the 5'-proximal structural elements of CoV gRNA transcripts, SL1 and SL5a/b/c in particular, were found to promote the formation of alternative conformations. Furthermore, we identified RNA-binding peptides from a pool derived from SARS-CoV N protein. These RNA-interacting peptides were shown to preferentially bind to wild-type SL5a RNA. In addition, our observations of N protein condensate formation *in vitro* demonstrated that liquid-liquid phase separation (LLPS) of N proteins with CoV-5'-UTR transcripts was influenced by the presence of SL5a/b/c. In conclusion, these results collectively reveal previously uncharacterized binding features between the 5'-proximal transcripts of CoV gRNAs and N proteins.

ARTICLE HISTORY

Revised 1 February 2025
Accepted 19 February 2025





KEYWORDS

N-protein; encapsidation; coronavirus; nucleocapsid; cis-acting elements; 5'-UTRs; RNA structure

1. Introduction

Since the emergence of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) in 2003, numerous studies have been conducted to understanding its life cycle. However, the molecular basis of nucleocapsid formation, a critical step in virion assembly, remains poorly understood [1–5]. In late 2019, a related virus, SARS-CoV-2, emerged and subsequently caused the global Coronavirus Disease 2019 (COVID-19) pandemic [6–8]. In coronaviruses (CoVs), nucleocapsids are mainly composed of genomic RNA (gRNA) and associated nucleocapsid (N) proteins [9]. It has been proposed that the encapsidation of gRNA is specifically initiated by the recognition of the so-called Packaging Signal (PS) that is exclusively possessed by viral gRNAs [10–12]. During this process, full-length gRNAs, but not the various sub-gRNAs (sgRNAs) or cellular mRNAs, are selectively encapsulated by N proteins and assembled into virus particles along with Membrane (M), Envelope (E), and Spike (S) proteins. Thus, the selective RNA-protein interaction between PSs and the N proteins is crucial for CoV genome packaging.

Structural models of SARS-CoV and SARS-CoV-2 N proteins have been partially resolved for their N-terminal domain (NTD) and C-terminal domain (CTD) by several independent research groups [13–19]. However, the crystal structures of the full-length N proteins remain unavailable, likely due to technical difficulties in structure studies. Nonetheless, the N protein is known to consist of several characteristic domains, including the N-arm, the RNA-binding NTD, the disordered Linker region (LKR) with the Serine (S)/Arginine (R)-rich motifs, the dimerization CTD, and the C-tail [16,20]. It has been suggested that the NTD plays a major role in RNA binding, while the CTD is primarily responsible for self-association and dimerization [21]. Additionally, the flanking sequences of the NTD and CTD, including the N-arm, LKR, and C-tail, have been proposed to be intrinsically disordered regions (IDRs). These regions contribute to the characteristic features of the N protein, such as its high positive charge and strong RNA-binding affinity [22–25]. Evidence also suggests that the RNA-binding affinity of the

CONTACT Yeh Chen  chyeah6599@nchu.edu.tw  Institute of Translational Medicine and New Drug Development, College of Medicine, China Medical University, Taichung City 40454, Taiwan; Chien-Hung Yu  chienhung_yu@mail.ncku.edu.tw  Department of Biochemistry and Molecular Biology, College of Medicine, National Cheng Kung University, Tainan City 70101, Taiwan

*Department of Food Science and Biotechnology, National Chung Hsing University, Taichung, 402202, Taiwan

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

N protein depends not only on the NTD but also on the CTD. This indicates that in addition to mediating oligomerization, the CTD may be involved in RNA binding and/or PS recognition [3,26,27]. Despite these findings, the detailed molecular basis of RNA binding by the N proteins is still poorly understood, though positively charged residues have been identified as crucial. Mapping the RNA-binding motifs in the N protein is therefore important for advancing our knowledge of CoV gRNA encapsidation, particularly from the perspective of selective packaging.

Our research group and others have previously proposed that PSs in beta-CoVs may likely be situated either in the 5'-proximal untranslated region (UTR) or the coding region of Non-structural protein 15 (Nsp15) [9–12,26,28–30]. However, the precise location and sequence of the PS in SARS-CoV-2 have not been identified, despite Terasaki et al. identifying a 1.4-kb-long sequence located in the nsp12 and nsp13 coding regions that can mediate efficient viral RNA packaging [31]. The secondary structures of these PSs have been found to be lineage-specific, often featuring multiple clustered structural motifs, such as the highly conserved 5'-gUUUCGUc-3' hexaloop motifs found in SARS-CoV and SARS-CoV-2 5'-UTR [32]. Recently, *in vitro*-synthesized recombinant transcripts and defective RNAs containing the 5'- and 3'-proximal sequences of SARS-CoV-2 gRNA were shown to be replicated and packaged with the helper viruses [33,34]. These findings support the hypothesis that the PS of SARS-CoV-2 is possibly located in the 5'-proximal regions of the gRNA, consistent with the observations for SARS-CoV and Transmissible gastroenteritis virus (TGEV) [9,28,35]. In this study, we characterized the binding features of SARS-CoV-2 5'-proximal gRNA transcripts and N proteins *in vitro* using primarily Electrophoretic Mobility Shift Assays (EMSAs) and attempted to identify the core RNA element(s) and peptide motif(s) crucial for their interaction.

2. Materials and methods

2.1. RNA synthesis by *in vitro* transcription

The cDNA sequences of CoV gRNAs were generated by reverse transcription of viral gRNAs using random-hexamer primers, including SARS-CoV-Frankfurt-1 (AY291315.1), SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512.2), Mouse hepatitis virus (MHV)-A59 (NC_048217.1), and Human CoV (HCoV)-229E (NC_002645.1). The 5'-proximal ~300-nt cDNA fragments were cloned into pUC19 vectors, while mutants of these constructs were generated with site-directed mutagenesis (Figure S2). The DNA templates for *in vitro* synthesis of RNA transcripts were generated by PCR using specific primer pairs (SARS-CoV-F/B: 5'-gaaattaatacagactcactataggtataggttttacc-taccaggaaaagc-3'/5'-ttctcgttgacaccaagaacaaggctc-3', SARS-CoV-2-F/B: 5'-gaaattaatacagactcactataggtataagaggtttatcctcc-cagg-3'/5'-tgttttctcgttgaaccagggaagagctctc-3', MHV-A59-F/B: 5'-gaaattaatacagactcactataggtataagagtgattggcgctccgtacgtac-3'/5'-atttgcccatctttgccattatgcaacctatgg-3', HCoV-229E-F/B: 5'-gaaattaatacagactcactataggtactaagcttatctatctacagatag-3'/5'-acggttgacggccattaggaacagttac-3', in which the forward primers have additional T7 promoter sequences. The

amplified DNA fragments were size-selected by electrophoresis with native agarose gels, followed by gel extraction and purification with phenol/chloroform. Purified T7-DNA templates were used to synthesize RNA transcripts with Amicon® Large Scale RNA Production System (Promega, USA), followed by treatment with RNase-free DNase I (Promega, USA), phenol/chloroform extraction, and ethanol precipitation. Fluorescently labelled SARS-CoV-2 SL5a and its mutant RNA oligos were synthesized by Integrated DNA Technologies (New Jersey, USA), with 5' 6-FAM (Fluorescein) and 5' Cy5™ fluorophores incorporated into the oligos, respectively. The purified 5'-transcripts and the fluorescence-labelled RNA oligos were subjected to EMSA with N protein and peptides, respectively.

2.2. Preparation of proteins and peptides

The production procedures of SARS-CoV-2 N proteins followed the methods previously reported by Ko et al. [36,37]. In brief, the full-length cDNA encoding SARS-CoV-2 N protein (isolate Wuhan-Hu-1, accession no. NC_045512) was synthesized chemically with codon optimization and cloned into the pET21 vector to create pET21-SARS-CoV-2-N for the expression of C-terminal His6-tagged recombinant N proteins. *Escherichia coli* strain OverExpress™ C43 (DE3) (Sigma-Aldrich, USA) was transformed with pET21-SARS-CoV-2-N and cultured in lysogeny broth (LB) medium at 37°C until OD600 reached 0.6–0.8. Overexpression of SARS-CoV-2 N proteins was induced by the addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) (Sigma-Aldrich, USA) into the culture media, followed by incubation at 15°C for 20 h. The cells were harvested by centrifugation, resuspended in lysis buffer consisting of 25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% glycerol, 1 mM tris(2-carboxyethyl) phosphine (TCEP) (Sigma-Aldrich, USA), and 1 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich, USA). Suspended cells were subjected to sonication on ice for lysis, followed by centrifugation at 28,000 g for 30 min at 4°C. The supernatant of the lysate was filtered with a 0.45 μm filter and loaded onto a HisTrap FF column (GE Healthcare, USA), followed by washing with lysis buffer containing 10 mM imidazole. The bound target proteins were eluted using a 20–200 mM imidazole gradient in the lysis buffer and analysed by SDS-PAGE for purity. Fractions containing the purified target proteins were pooled and concentrated using Amicon® Ultra-15 Centrifugal Filter Units (10 kDa molecular weight cut-off) (Millipore, USA), and the concentrated protein solution was dialysed against storage buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 5% glycerol). For protein condensate observation, RNA transcripts and N proteins were mixed in RNase-free buffer (10 mM MgCl₂, 50 mM HEPES, pH 7.0) prior to microscopic observation. The concentrations of NaCl, protein, and RNA which have been suggested as key to the formation of N-protein condensates were optimized as shown in Figure S4. The peptide library, consisting of 80 15-a.a. peptides spanning the whole SARS-CoV N protein (Urbani strain), was chemically synthesized, followed by HPLC purification (Kelowna International Scientific Inc., Taiwan). The sequences of the entire peptide library are listed in

Table S1. Purified proteins and peptides were subjected to EMSA with RNA transcripts, respectively.

2.3. EMSA of RNA transcripts with N proteins and peptides

For EMSAs, synthesized and purified RNA transcripts of the 5'-proximal sequences of different CoV gRNAs and their mutants were subjected to refolding in the buffer containing 50 mM HEPES (pH 7.5), 20 mM NaCl, and 10 mM $MgCl_2$ at 85°C for 30 s, followed by cooling at ambient temperature, except where otherwise specified. In general, 50 ng RNA transcripts were tested for binding to serially diluted N protein and peptides, respectively, by incubating the RNA-protein mixture at 4°C for 20 min and subjected to EMSA with native PAGE (4% 1:19 acrylamide) in a MOPS buffer system (200 mM MOPS, 50 mM NaOAc, pH 7.5). The native gels were stained with SYBR[™] Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific Inc., USA) to visualize the mobility shifts of 5'-proximal gRNA transcripts upon binding to the N proteins and peptides, respectively. For EMSA with RNA oligos, FAM-SL5a-Wt and Cy5-SL5a-Mut oligos were mixed and incubated with serially diluted peptides and N proteins, respectively, followed by native PAGE. The gel was imaged using an Amersham[™] Typhoon[™] Biomolecular Imager (GE Healthcare, USA) with detection filters specific for Fluorescein and Cy5, respectively, to distinguish the band shifts of 5'-FAM-SL5a-Wt and 5'-Cy5-SL5a-Mut oligos.

3. Results

3.1. Deletions of SL1 or SL5a/b/c alter the conformations of SARS-CoV-2 5'-proximal transcripts of gRNAs

We have previously reported that the secondary structure of the 5'-proximal regions of CoV gRNAs exhibits a lineage-specific feature with five conserved structural elements, designated Stem-loop (SL) 1 to 5 [28,32]. The canonical secondary structure of the SARS-CoV-2 5'-UTR is shown in Figure S1A [38–44], which is highly similar to what has been reported for SARS-CoV and other related CoVs in the sub-genus Sarbecovirus [28,32]. To verify whether the 5'-proximal sequence of SARS-CoV-2 gRNA has a stable major conformation and if the structural elements SL1–5 contribute to its conformational stability, we studied the

electrophoretic mobilities of in vitro synthesized 5'-proximal gRNA transcripts with differential SL deletions. The secondary structures and the primary sequences of the studied SARS-CoV-2 and HCoV-229E transcripts are shown in Figure S1 and S2. It was shown that the wild-type 5' RNA transcripts have a major band in native gels (Figure 1). However, in the presence of 20 mM magnesium ions (Mg^{2+}), additional upper bands appeared, indicating that alternative conformers with significantly lower mobility formed under such conditions. Notably, transcripts with an SL1 deletion exhibited a significantly increased proportion of alternative conformers in the presence of as little as 5 mM Mg^{2+} (Figure 1), indicating that the equilibrium between the major and alternative conformations of SL1-deleted transcripts is relatively more sensitive to Mg^{2+} , which is widely reported for its crucial role in stabilizing RNA structures and promoting functional activities [45].

We observed a more significant conformational shift for the transcripts with deleted SL5a/b/c (Figure 1), in which the highly conserved 5'-gUUUCGUc-3' structural motifs are apically located [32]. Other deletion mutants that may interfere with the formations of SL5 and/or the apical structural motifs, such as the transcript $\Delta 5'SL5$ in which the 5'-side sequences of the SL5 stem region are deleted, also showed modest Mg^{2+} -dependent conformational shifts (Figure 1), while transcripts $\Delta SL2/3$ and $\Delta SL4$ have the least conformational shifts. It has been speculated that the apical structural elements SL5a/b/c, present in the SL5s of alpha- and beta-CoVs, promote selective packaging [28,32]. Here we found that SL5a/b/c play critical roles in the conformational stability of SARS-CoV-2 5'-proximal gRNA. To verify whether the SL conformation represents the major form of the 5'-proximal transcripts in native gels, we generated mutant Rev1, in which multiple base-pairing nucleotides were swapped in the stem regions of SL1–5 to eliminate potential alternative conformations while maintaining the SL conformation as shown in Figure S1. Figure 2A shows that the major bands of wild-type and the mutant transcripts have comparable mobility in native gels, revealing their conformational similarity. Figure 2B shows that Rev1 mutants exhibit no observable conformational shifts in the presence of Mg^{2+} , unlike wild-type and SL-deleted transcripts (Figure 1), suggesting that the nucleotide substitutions in Rev1 transcripts relatively stabilize SL-conformations.

To investigate whether other CoV 5'-proximal sequences also undergo Mg^{2+} -dependent conformational shifts, we

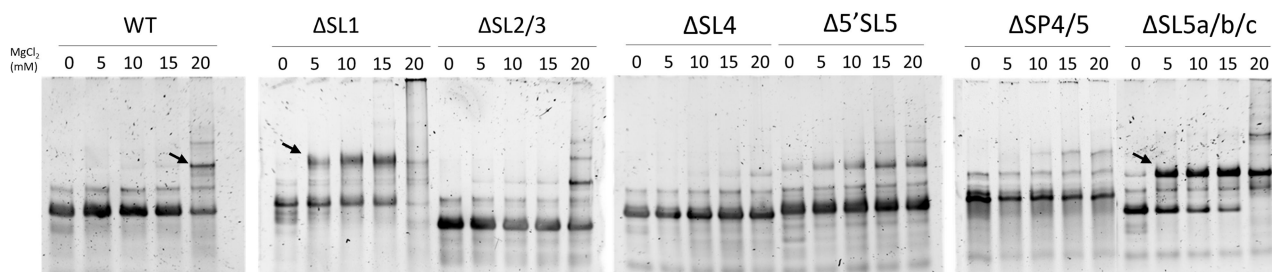


Figure 1. EMSA of SARS-CoV-2 5'-proximal gRNA transcripts. The wild-type (WT) and mutants with deleted SL1 ($\Delta SL1$), SL2/3 ($\Delta SL2/3$), SL4 ($\Delta SL4$), 5'-sequence of SL5 ($\Delta 5'SL5$), spacer sequence between SL4 and 5 ($\Delta SP4/5$), and SL5a/b/c ($\Delta SL5a/b/c$) were analysed for mobility shifts in native acrylamide gels. Each transcript was refolded prior to electrophoresis in the presence of varying concentrations of $MgCl_2$, as indicated. The arrows highlight the major $MgCl_2$ -dependent conformationally shifted transcripts with significantly altered mobility. The sequences of these 5'-transcripts are provided in Figure S1 and S2.

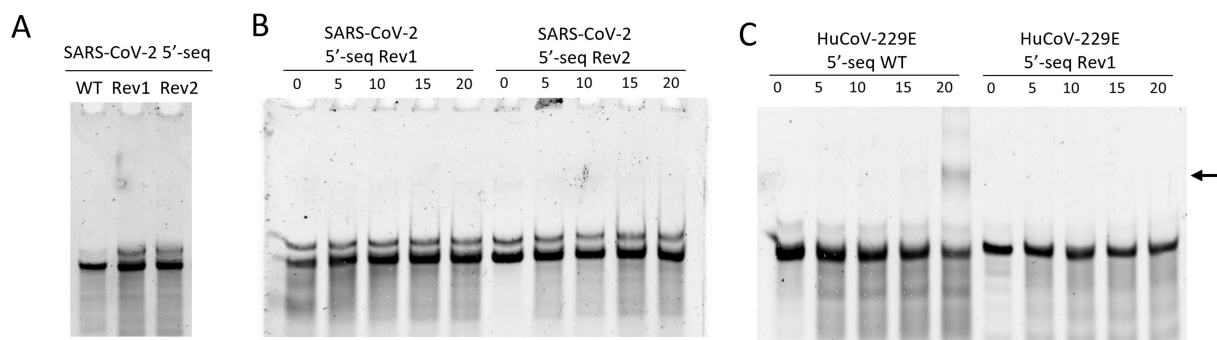


Figure 2. EMSA of mutant CoV 5' RNA transcripts. (A) Synthesized wild-type and the mutant 5'-transcripts of SARS-CoV-2 Rev1 exhibited similar major bands in native acrylamide gel. Transcripts of (B) SARS-CoV-2 Rev1 and (C) HCoV-229E wild-type (WT) and Rev1 were refolded in differential concentrations of MgCl₂ prior to native acrylamide gel electrophoresis, respectively, in which multiple mutations were applied to abolish potential alternative conformations while the SL-conformation was maintained. The arrow indicates the positions of conformational shifted transcripts with significantly changed mobility. Sequences of HCoV-229E transcripts were listed in Figure S2.

generated wild-type and Rev1 mutant transcripts of HCoV-229E (Figure S1 and 2). It was revealed that the wild-type transcripts exhibit alternative conformations in the presence of 15–20 mM Mg²⁺ but not the mutated transcript Rev1 (Figure 2C). These data suggest that the native 5'-proximal transcripts of SARS-CoV-2 and HCoV-229E may shift from a major conformation to alternative conformations in the presence of Mg²⁺, and deletions of SL5a/b/c result in more significant conformational shifts.

3.2. SARS-CoV-2 N protein preferentially binds to SL-conformer

Previously, the PS of TGEV was proposed to reside in the 5'-proximal sequences of gRNA [35]. Recently, it has been found that the 5'-proximal sequence of SARS-CoV-2 gRNA also participates in selective packaging [33]. Here, we verified the binding between SARS-CoV-2 N proteins and the 5'-proximal transcripts of gRNAs using EMSA. The results show that the binding of in vitro synthesized RNA transcripts of the SARS-CoV-2 5'-proximal gRNA is slightly stronger than that of non-specific RNA transcripts derived from Firefly luciferase (Fluc) coding sequences. This is evident from the reduced presence of unbound free RNAs, although the bound complex bands appear at lane 5 (Figure 3). This result suggests that the N protein preferentially interacts with structured RNA transcripts. However, the abundant positive charges of the N protein also facilitate non-specific binding to a variety of RNAs to a certain extent [46]. We also found that the 3'-transcripts of SARS-CoV-2 gRNA exhibit relatively high binding affinity to the N proteins (Figure 3), consistent with

previous reports that N protein promotes replication, transcription, and/or translation by interacting with the highly structured 3'-end of gRNA [39–41]. The strong binding between the 3'-end and the N protein may facilitate selective packaging synergistically with the 5'-sequence [33]. However, the critical PS is unlikely to reside in the 3'-region because all the sgRNAs also possess 3'UTRs as gRNA does; yet only the full-length gRNAs are predominantly packaged into virions. Therefore, the 5'-proximal region is more likely to play a critical role in directing selective genome packaging and warrants further investigation.

In the comparison of the seven CoVs prevalent among humans (Figure 4), HCoV-OC43 and HCoV-HKU1 exhibit sequence insertions that fold into structural elements with repetitive structural motifs in the Nsp15 coding sequence, similar to the characteristic MHV-PS. In contrast, the other five HCoVs, HCoV-229E, HCoV-NL63, Middle East Respiratory Syndrome-related CoV (MERS-CoV), SARS-CoV, and SARS-CoV-2 lack these sequence insertions and homologous structures (Figure 4A). Instead, these non-Embecovirus human beta-CoVs possess additional conserved structural elements, SL5a, -b, and -c, in the 5'-proximal region of their gRNAs (Figure 4B). This feature suggests that SL5a/b/c in SARS-CoV-2 gRNA may functionally correspond to Embecovirus PS for encapsidation (Figure 4C,D).

To investigate whether stabilizing the SL-conformation and eliminating alternative conformations of the 5'-proximal gRNA transcripts facilitates N protein binding, mutated transcripts were subjected to EMSA and compared with wild-type transcripts. Figure 5 shows that the SARS-CoV-2 mutant Rev1 exhibits a stronger association with the N proteins, evidenced

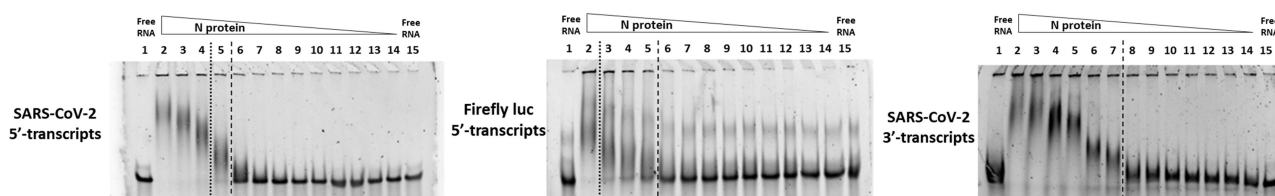


Figure 3. EMSA of SARS-CoV-2 5'-proximal gRNA transcripts with N proteins. The ~300-nt transcripts of SARS-CoV-2 5'-proximal gRNA, firefly luciferase 5'-proximal coding sequences, and SARS-CoV-2 3'-proximal gRNA sequences were subjected to EMSA with serially diluted SARS-CoV-2 N proteins. The initial molar ratio between RNA transcripts and N proteins was 1:6.6, followed by 3/4X serial dilutions. Dashed lines indicate the lanes exhibiting partial or complete band shifts corresponding to the RNA-N protein complexes, while the dotted line marks the lanes showing partially or fully unbound free RNA transcripts.

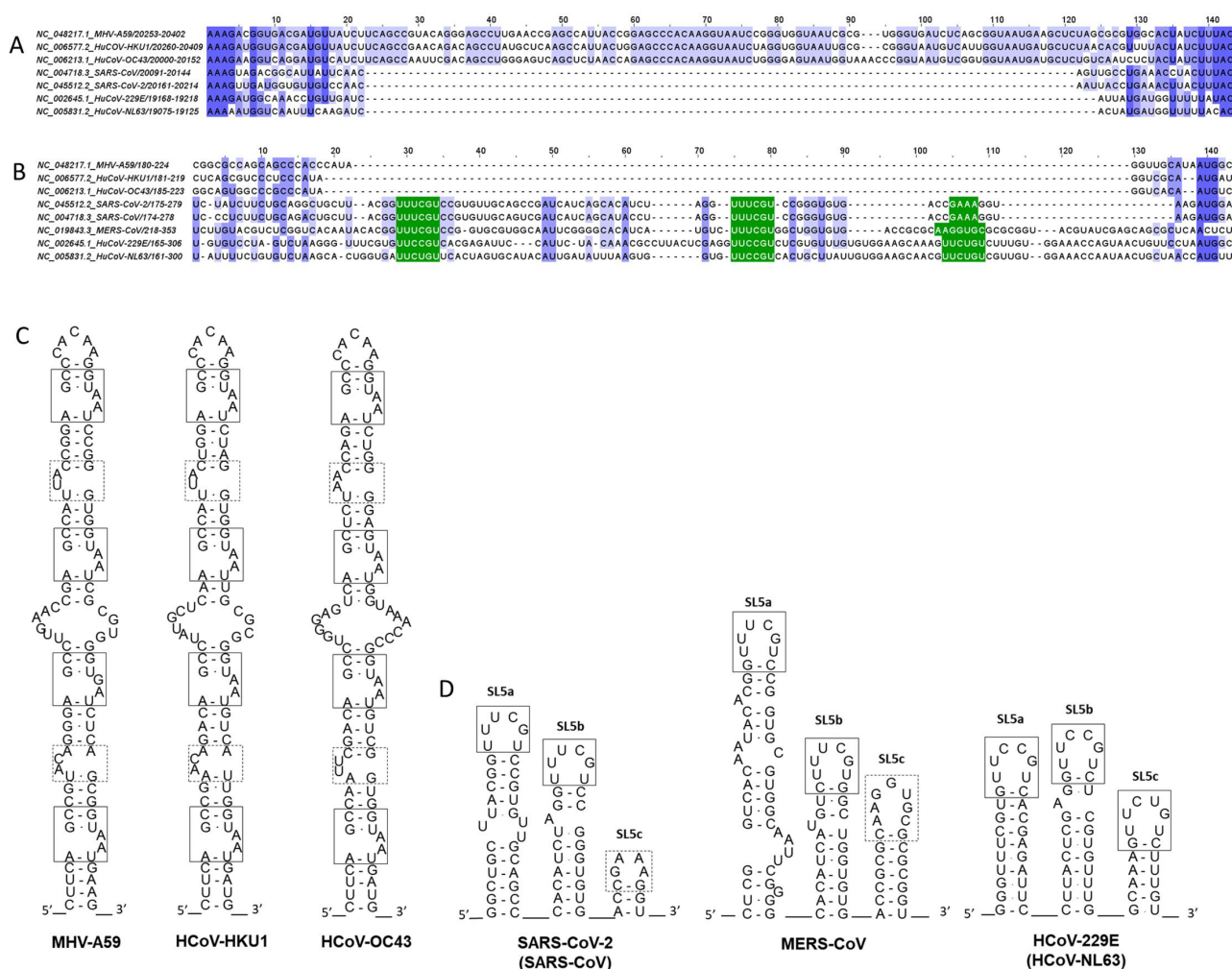


Figure 4. Potential counterparts of the conserved ps-like structural elements identified in gRNAs of human CoVs. (a) Multiple sequence alignments of regions corresponding to the Nsp15-located embecovirus-ps and (b) the 5'UTR-located 5'-UUUYCGU-3' structural motifs (highlighted in green) are shown for the model CoV MHV and various human CoVs. (c) Predicted secondary structure models of the MHV-PS and its potential counterparts in HCoV-HKU1 and HCoV-OC43 are displayed. (d) Secondary structure predictions of clustered conserved motifs in the 5'UTRs of non-embecovirus human CoVs are shown. Highly conserved repetitive structural motifs are enclosed with solid-line boxes, while less conserved motifs are outlined with dashed-line boxes.

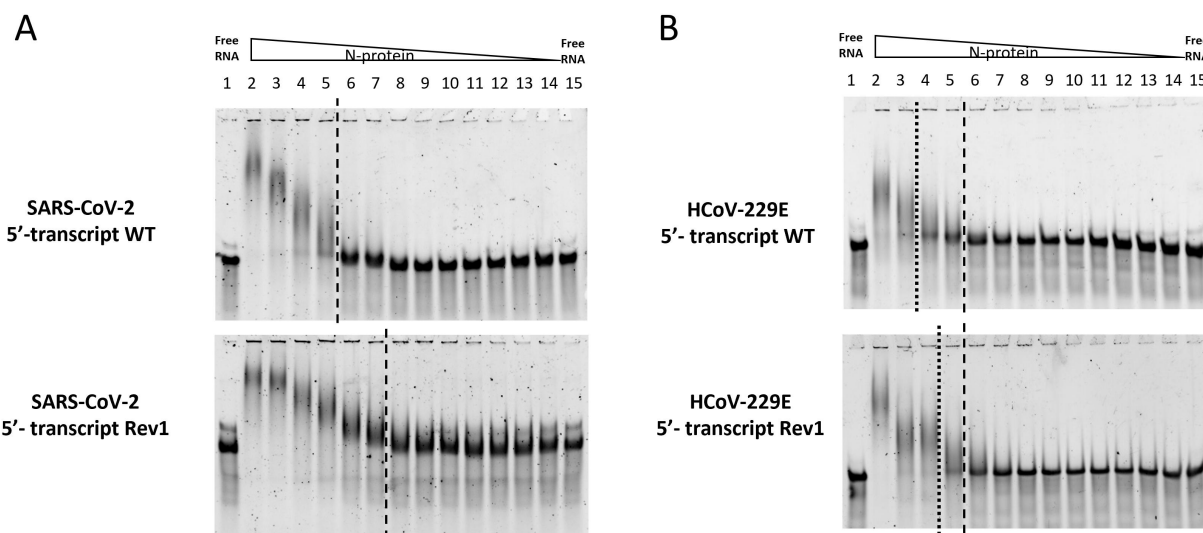


Figure 5. EMSA of 5'-proximal gRNA transcripts with N proteins. Wild-type and Rev1 mutants of (a) SARS-CoV-2 and (b) HCoV-229E 5'-transcripts were tested for their binding to serial diluted SARS-CoV-2 N proteins. The molar ratio between RNA transcripts and N proteins started from 1:6.6 with 3/4X serial dilutions. Sequences of these mutant transcripts were listed in Figure S1 and 2. Dashed lines indicate lanes showing (partially) shifted bands of RNA-N protein complexes, while dotted lines mark lanes with (partially) unbound free RNA transcripts.

by more significant band shifts at relatively lower concentrations of N proteins. These results indicate that sequence alterations do not abolish the association between the 5'-transcripts and the N protein as long as the SL-like conformation is maintained. Moreover, the relatively weaker binding shown by the wild-type 5'-transcripts suggests that the native sequence is less stable than the mutant for N protein binding. A similar trend was observed in the binding between 5'-transcripts of HCoV-229E Rev1 and the N protein, which showed relatively stronger binding compared to the wild-type (Figure 5B). This finding suggests that the native 5'-proximal transcripts of CoV gRNAs have a relatively more flexible structural feature with potentially less rigid SL-like conformers for N protein binding. One possible benefit of this flexibility is that the 5'-proximal region, which contains multiple cis-acting elements, may undergo conformational switches to facilitate its diverse regulatory roles in the viral life cycle, including replication and translation [47–49]. Such conformational switching upon N protein binding may be critical for the N protein-mediated regulation of viral replication, transcription, and translation [50–54].

3.3. Identification of N protein-derived peptides that bind to CoV 5'-proximal transcripts and SL5a oligos

The NTD of SARS-CoV N protein has been implicated in RNA binding [3,14,55], while in other CoVs, the CTD has also been shown to possess RNA-binding ability [19,56–58]. However, the critical peptide motifs within the N protein responsible for RNA binding have not been systematically investigated. To address this, we generated a synthetic library consisting of eighty 15-a.a. peptides spanning the entire SARS-CoV N protein (Table S1). This peptide library was subjected to RNA-binding assays with the transcripts of the SARS-CoV 5'-UTR and HCoV-229E SL5, respectively, to identify strong RNA binders. Figure 6A shows

the EMSA results of the RNA transcripts tested with 16 subsets of pooled peptides in the first round of screening. Peptide pools that caused significant band shifts were further analysed using individual peptides (Figure 6B). Ultimately, five peptides with significant RNA-binding ability were identified as listed in Table S1, designated as Peptide #7, #8, #17, #37, and #51.

Mapping these peptide sequences onto structural models of SARS-CoV-2 N protein revealed their locations: Peptides #7 and #8 are in the intrinsically disordered N-arm; Peptide #17 spans the $\alpha 1$ and $\beta 2$ motifs of the NTD; Peptide #37 is located at the centre of the S/R-rich region; and Peptide #51 corresponds to the $\alpha 0$ - $\eta 1$ motifs at the N-terminus of the CTD [3,4,16,25]. A common feature among these peptides is their positive charge (Table S1). However, the presence of positive charges alone does not fully explain their RNA-binding activities, as these peptides were not the most positively charged peptides among all the peptides in the library. Some other highly positively charged peptides, such as Peptides #52 and #73, were not capable of binding RNAs (Table S1). Another common characteristic of the RNA-binding peptides is the presence of R residues. However, peptides with high R content, such as #18 and #38, did not show significant RNA binding. Intriguingly, recent studies reported that R92E and R107E mutations significantly reduce the RNA-binding ability of SARS-CoV-2 N protein [25], highlighting the importance of these conserved R residues. These two specific R residues map to peptides #17–19 and #20–22, yet only Peptide #17 exhibited significant binding to the RNA transcripts. Collectively, these findings indicate that both sequence and structural contexts are critical for these peptides to interact with CoV RNA transcripts, even though positive charges and R residues are hallmarks of RNA-binding peptides.

We further tested whether these SARS-CoV N protein-derived peptides, particularly those mapped to NTD, could bind to 5'-transcripts of other CoVs. Figure 7 shows the EMSA

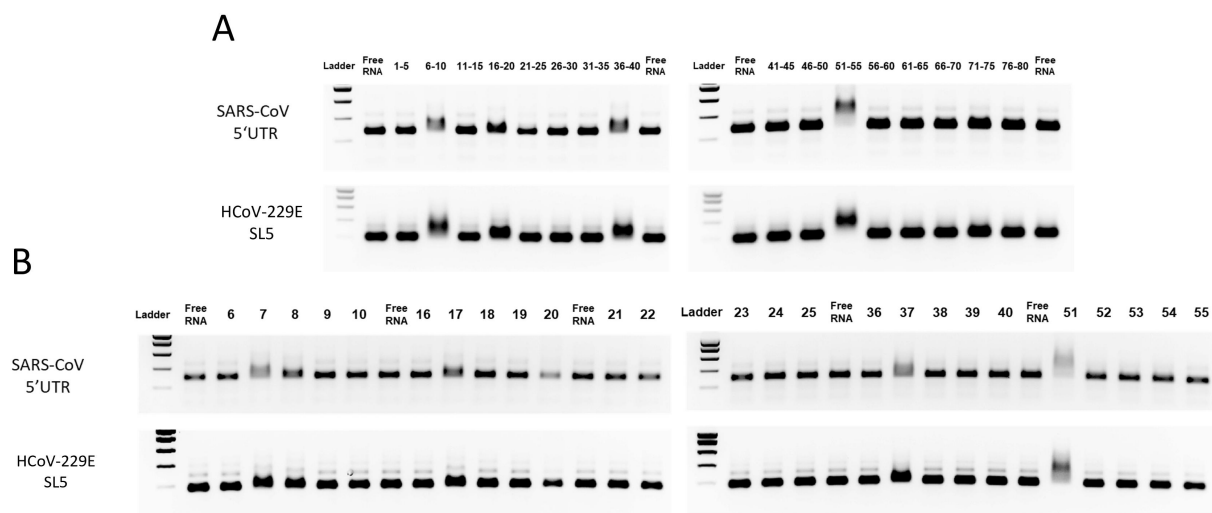


Figure 6. Screening of RNA-binding peptides from the N protein-derived peptide library. Transcripts of SARS-CoV 5'UTR and HCoV-229E SL5 were subjected to EMSA with (a) pooled peptides and (b) individual peptides derived from SARS-CoV N protein.

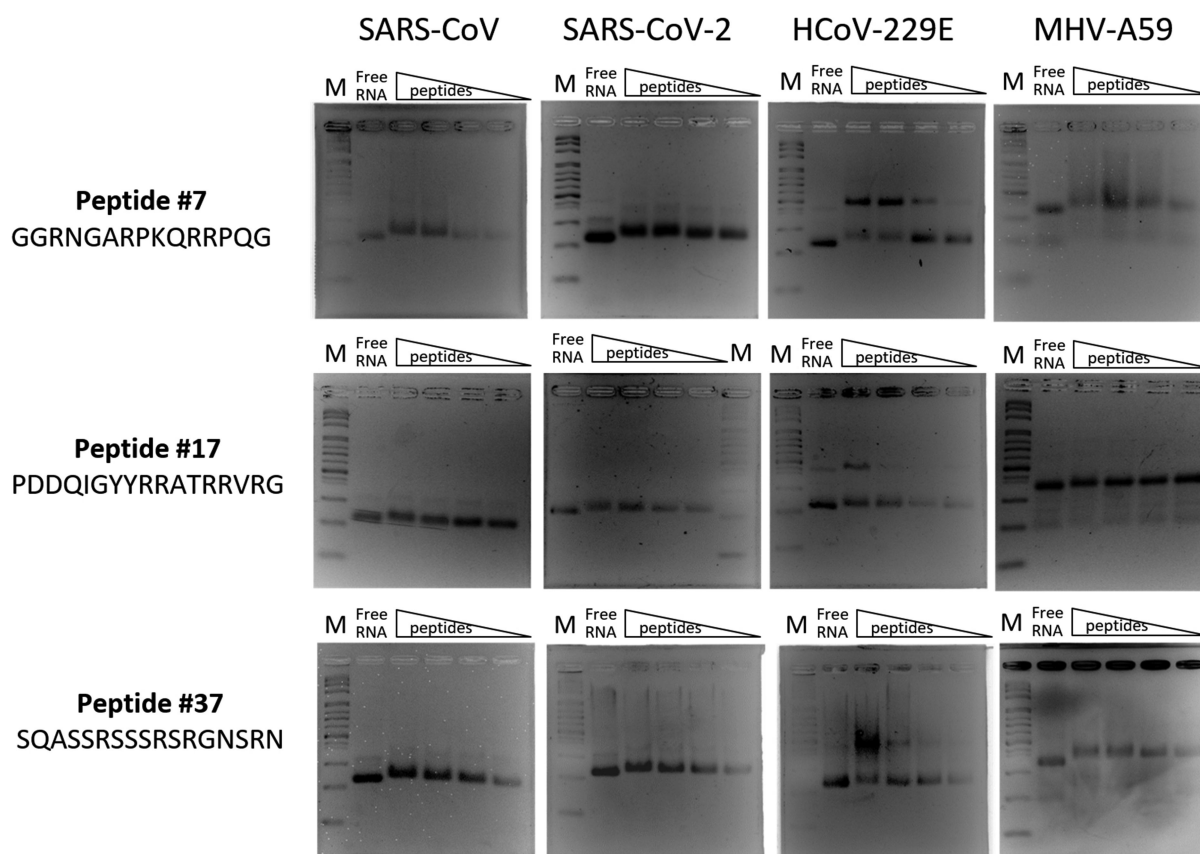


Figure 7. EMSA showing interactions between CoV 5'-proximal RNA transcripts and the N protein-derived peptides. The 5'-proximal gRNA transcripts of SARS-CoV, SARS-CoV-2, HCoV-229E, and MHV-A59 were subjected to EMSA with the peptide #7, #17, and #37, respectively. Each lane contains 1000 ng of RNA transcripts with 1/2X serial diluted peptides starting from 2 ng/ μ l in a 10- μ l mixture.

results for Peptides #7, #17, and #37 with the 5'-proximal transcripts of SARS-CoV, SARS-CoV-2, HCoV-229E, and MHV-A59, respectively. These peptides interacted with all tested transcripts without significant differences under our experimental conditions. Notably, peptide-dependent conformational shifts were observed in the 5'-proximal gRNA transcripts, particularly for HCoV-229E. A slower-migrating band was detected for HCoV-229E 5'-transcripts in the presence of peptides (Figure 7), with mobility similar to MHV-A59 5'-transcripts, which lack the SL5a/b/c region (Figure 7). This finding suggests that SL5a/b/c contributes to forming a more compact SL-conformation, consistent with our comparison of WT and Δ SL5a/b/c transcripts of SARS-CoV-2 (Figure 1). Furthermore, these data suggest that certain peptides may facilitate conformational changes, as N proteins have been reported for their chaperone activity [59].

To investigate whether N protein and its derived peptides preferentially bind the apical 5'-gUUUCGUc-3' motif of SL5a/b, we synthesized RNA oligos resembling this motif and a non-specific hexa-loop 5'-gAACAAAc-3'. Figure 8A shows the loop sequences of the 5'-FAM-labelled SL5a-Wt and the 5'-Cy5-labelled mutant RNAs. Equal amounts of these two differentially labelled RNA hairpins were pre-mixed and subjected to competition EMSA with N protein (Figure 8b) and Peptide #17 and #37 (Figure 8C,D), respectively. Distinct green and red fluorescence channels visualized SL5a-Wt and SL5a-Mut oligos, respectively. The results show that fewer wild-type probes

(green) remain unbound compared to mutant probes (red) at the same N protein concentration (Figure 8B), indicating a binding advantage for SL5a-Wt. Similarly, SL5a-Wt showed stronger binding to Peptide #17 (Figure 8C) but not #37 (Figure 8D). These findings suggest that the loop sequence influences N protein binding. One potential mechanism is that the G-residue in the motif is recognized by the G-specific binding pocket of the N protein [60]. However, this G-binding pocket has been reported to reside in the CTD, whereas Peptide #17, which showed differential binding, is in the NTD. These results suggest that the selective binding of the N protein to SL5a may be synergistically facilitated by multiple RNA-recognizing motifs. Further research is required to explore this mechanism.

3.4. The formation of N-protein condensates *in vitro* with CoV 5'-proximal transcripts

It has been demonstrated that purified SARS-CoV-2 N-proteins undergo LLPS *in vitro* to form droplets in solution [61–63]. This process is sensitive to salt concentration, temperature, protein concentration, and the presence of RNA. In our experiments, we observed that SARS-CoV-2 N-proteins form more droplets in the presence of 200 mM NaCl compared to low (20 mM) or high (800 mM) NaCl concentrations (Figure S3A). Additionally, we found that fluorescence-labelled RNA probes co-localize with these droplets (Figure S4A). Previous studies have suggested that

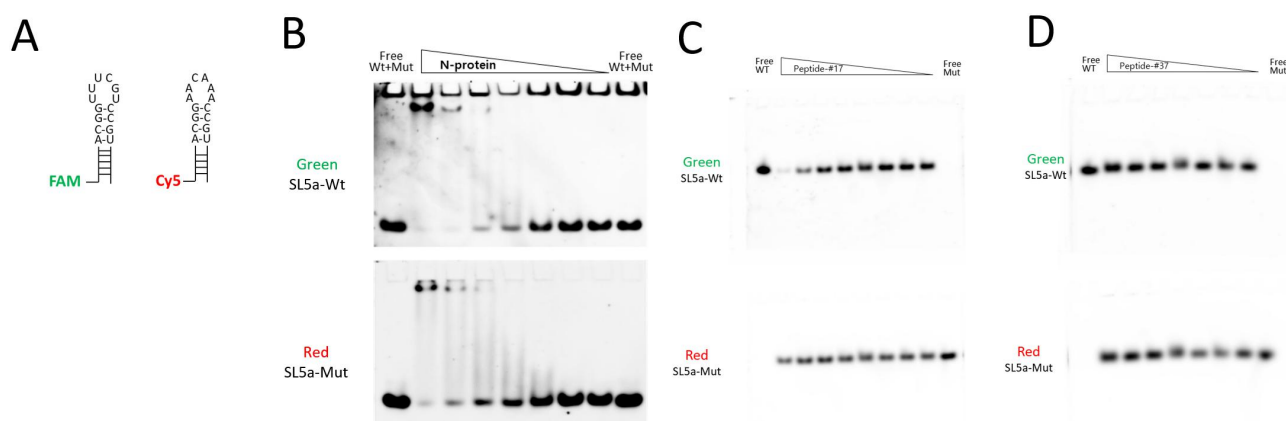


Figure 8. The binding between SL5a RNA oligos and N proteins/peptides. (A) Fluorescein (fam)-labelled wild-type SL5a RNA oligos, containing the native 5'-gUuucguc-3' hexa-loop motif, and Cy5-labelled mutant SL5a RNA oligos, containing the altered 5'-gAcaaac-3' loop sequence, were subjected to competition EMSA. (B) EMSA analysis of SL5a RNA oligos with SARS-CoV-2 N protein. Each lane contains 28.4 nM of fam-labelled wild-type SL5a and Cy5-labelled mutant SL5a, incubated with serially diluted N proteins (1/2X dilutions), starting from 85 ng/μl. The green fluorescence channel detected the fam-labelled wild-type oligos, while the red fluorescence channel detected the Cy5-labelled mutant oligos. (C-D) competition EMSA of SL5a wild-type and mutant RNA oligos (0.8 μm each) with serially diluted (starting from 10 ng/μl, 1/2X dilutions) peptides (C) #17 and (D) #37, respectively. The assays were performed using native PAGE, followed by fluorescence imaging.

SARS-CoV-2 gRNAs promote the LLPS of N-proteins [63–65]. Consistent with these findings, we observed condensate formation when SARS-CoV-2 N-proteins were incubated with their 5'RNAs (Figure 9). However, we found that an excessive amount of RNA transcripts abolishes droplet formation, leading to dense precipitation instead (Figure 9A). We also tested the condensate formation of MHV 5'RNAs with SARS-CoV-2 N proteins. By comparing wild-type and mutant MHV 5'RNA transcripts (Figure 9B), we found notable differences. The wild-type transcripts contain the typical *Embecovirus*-specific SL5 structure, which lacks SL5a/b/c, whereas the mutant transcripts were engineered to include *Sarbecovirus*-like SL5a/b/c (Figure 9C,D). Our observations revealed that the insertion of additional SL5a/b/c into the MHV 5'-transcripts significantly enhances condensate formation (Figure 9B). This suggests that SL5a/b/c may positively influence condensate formation. These *in vitro* LLPS experiments provide insights into the potential role of SL5a, b, and c in N-protein interactions and possibly genome packaging. Future studies using reverse-genetically engineered viruses are required to provide further evidence of these findings *in vivo*.

4. Discussion

The CoV N proteins are structurally heterogeneous RNA-binding proteins with multiple structural domains and IDRs [5,20,64,67–69]. It has been suggested that the SARS-CoV-2 N protein consists of two independently folded NTD and CTD which are connected by the S/R-rich LKR. In addition, two IDRs are found at the termini of both the NTD and CTD. It is generally known that the NTD is responsible for RNA binding, while the CTD is predominantly responsible for dimerization [25,56,58,70–74]. In addition to the broadly studied NTD, we demonstrated RNA-interacting activities for the IDRs in N proteins by screening the peptide library derived from N protein, which had been previously reported but less characterized [75]. Among the five RNA-binding peptides we have identified in this study, three were mapped to the IDRs, including Peptide #7 and #8 derived from the intrinsically disordered N-arm and Peptide #37 from the S/R-rich LKR. Recent

studies have shown that deletion of the N-arm decreases the range of RNA concentrations inducing phase separation of SARS-CoV-2 N protein [76], and the S/R-rich region has been suggested to have RNA-binding capacity under hydrogen – deuterium exchange mass spectrometry analysis [77]. These findings collectively suggest that the IDRs in CoV N proteins play crucial roles in regulating the RNA-binding and/or oligomerization dynamics upon engaging RNAs, as has been recently proposed [22–24,68,76,78–82]. An intriguing finding in this study was that only Peptide #37 was found to have strong RNA-binding activity, whereas other neighbouring peptides, e.g. #36, #38–40, showed no significant binding despite their high contents of S/R residues. This result indicates that other uncharacterized sequence and/or structure contexts could be crucial for RNA binding. Previous studies have reported that the S/R residues in the LKR are potentially phosphorylated/methylated and such modifications may regulate the binding of N proteins to the 5'UTRs of SARS-CoV and SARS-CoV-2 gRNAs [52,83,84]. However, the peptides tested in this study consisted of native a.a. without modifications. How phosphorylation and/or methylation may affect the RNA binding activity of S/R-rich peptides remains to be determined in the future. (why not #51)

The interactions between 5'-proximal transcripts of CoV gRNAs and N proteins have been shown to be crucial not only for genome packaging but also for replication. It was reported that MHV N protein binds to the transcriptional regulatory sequence (TRS) in the 5'-proximal sequences of gRNAs with high affinity and specificity, and such interaction plays an important role in discontinuous RNA transcription [50,51,53,85]. However, such TRS-binding specificity was found not significant in SARS-CoV [85]. Thus, it is not clear whether such specific binding of N protein to TRS, which is critical for efficient sgRNA synthesis of MHV, applies to all CoVs [54]. Intriguingly, these previous studies in MHV have revealed the helix unwinding properties of the N proteins that promote template switching. In TGEV and SARS-CoV, their N proteins were found to possess RNA chaperone activities which could

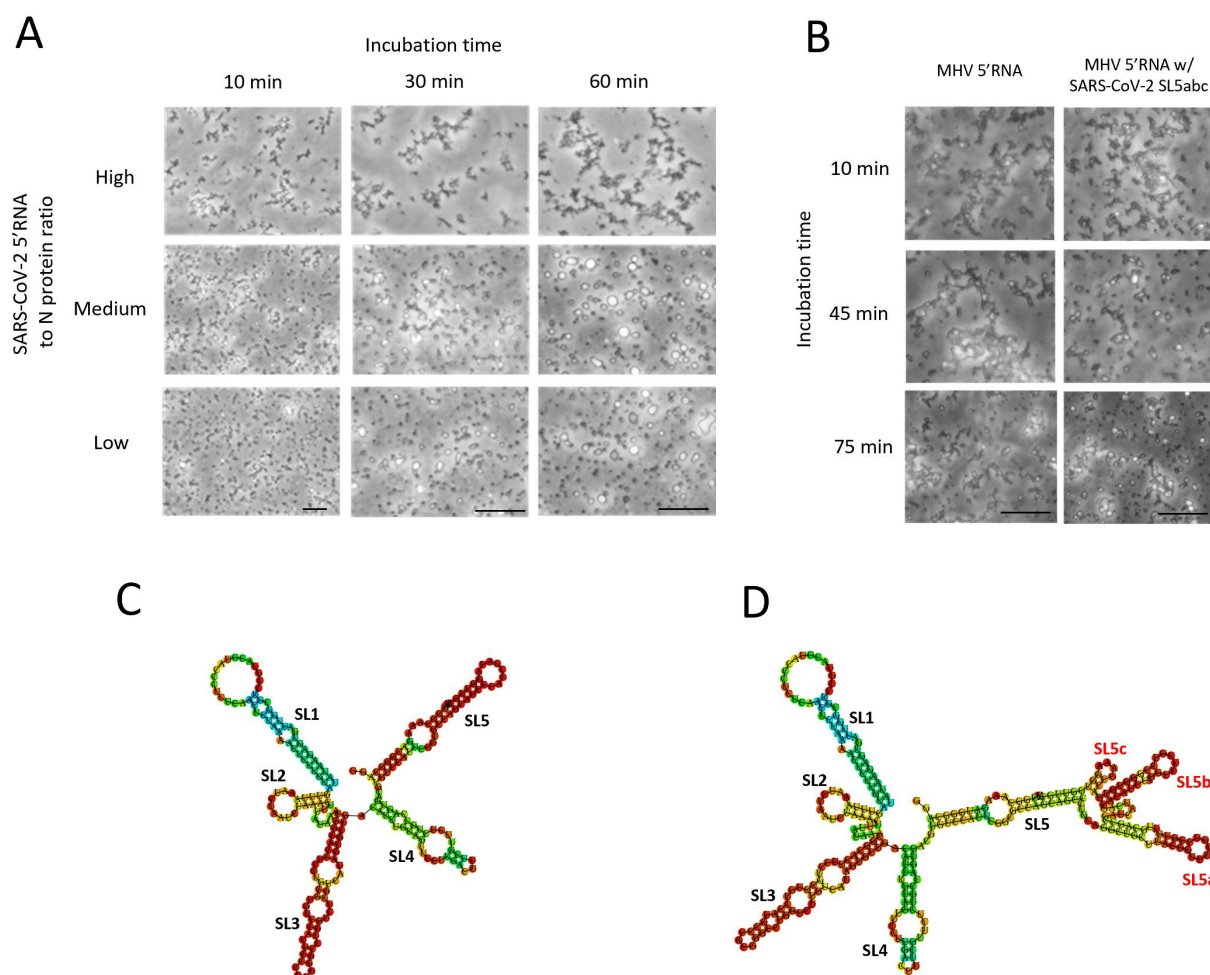


Figure 9. The N-protein condensates with SARS-CoV-2 and MHV 5'-proximal gRNA transcripts. Microscopic images show the formation of condensates between N proteins and in vitro synthesized 5'-proximal gRNA transcripts. (A) Condensates formed by SARS-CoV-2 N protein with SARS-CoV-2 5'-proximal gRNA transcripts were observed after 10, 30, and 60 minutes of incubation. The protein-to-RNA weight ratios tested were 1:0.9 (high), 1:0.05 (medium), and 1:0.005 (low), respectively. (B) Condensates formed by SARS-CoV-2 N protein with (C) wild-type and (D) mutant MHV 5'-proximal gRNA transcripts. The wild-type MHV 5'RNA lacks the apical SL5a, SL5b, and SL5c structural elements, while the mutant MHV 5'RNA is engineered to include these elements, mimicking the SL5a/b/c structures of SARS-CoV-2. The secondary structures of the wild-type and mutant MHV 5'RNAs were predicted using RNAfold [66]. Scale bar: 100 μ m.

modulate the conformations of RNAs [59]. Here we reported that Peptide #7, #17, and #37 have general RNA binding activities to 5'-proximal transcripts of four different CoVs (Figure 7), yet the peptide-induced conformational shifts were more significantly observed in some particular peptide-RNA pairs, e.g. Peptide #7 with HCoV-229E 5'-transcripts. This result indicates that 5'-transcripts of different CoV gRNAs have differential sensitivities to peptide-induced conformational shifts while different peptides may also have differential chaperone activities. In experiments probing the binding specificities of the peptides to wild-type SL5a over the mutant, Peptide #37 did not show significant selectivity. Peptide #17 showed preferential binding to the wild-type SL5a, which is mapped to the α 1 and β 2 motifs of the NTD. Possibly, selective packaging is initiated by specific recognitions mediated by NTD, while the IDRs may facilitate encapsidation with their non-specific RNA interacting ability, supporting the dynamic and flexible RNA binding models recently proposed by several research groups [22,24,68,81,82,86].

We showed in this study that the 5'-proximal structural elements, SL5a/b/c in particular, are crucial to promote the SL-conformation of the 5'-proximal transcripts of SARS-CoV-2 gRNAs for N protein binding. We also found that the N protein preferentially interacts with SL5a with the highly conserved 5'-gUUUCGUc-3' motif. These data support the idea that SL5 and its apical structural elements, SL5a/b/c, participate in selective genome packaging [9,28,32], in accordance with what has been reported for TGEV and SARS-CoV-2 that the 5'-proximal regions are critical for CoV genome packaging. Nonetheless, other regions in CoV gRNAs have also been reported to direct RNA packaging [12,26,87], though those tested regions lack specific structural and phylogenetic homology. For instance, Syed et al., showed that the segment stretching from nt 20,080 to 22,222 of SARS-CoV-2 gRNA promotes non-viral reporter mRNAs to be packaged into virus-like particles made of recombinant N, M, E, and S proteins [87]. This particular packaging-promoting segment spans across the coding sequences of Nsp15 and 16 and partial S protein. However, further

dissections of such ~2000-nt segment did not lead to the identification of specific elements similar to *Embecovirus* PS or *Sarbecovirus* SL5a/b/c. It was suggested that this particular segment has high proportions of double-stranded regions with non-specific structures. Terasaki et al. identified a 1.4-kb sequence within the coding regions of nsp12 and nsp13 in SARS-CoV-2 genomic RNA as critical for efficient viral RNA packaging [31]. While several structural elements are computationally predicted within this region, they do not exhibit significant structural homology or phylogenetic conservation across related coronaviruses. This lack of conservation raises questions about the specific mechanisms through which this sequence facilitates RNA packaging. It is possible that the packaging efficiency mediated by this region is not driven by conserved structural motifs but rather by a combination of its intrinsic properties, such as high proportions of double-stranded regions, sequence composition, or the dynamic interplay between RNA and viral proteins. Previously, double-stranded RNAs have been reported to have stronger interactions with N proteins than single-stranded RNAs [25,88,89,9091]. This may explain how this particular segment of gRNA was capable of promoting the formation of virus-like particles without possessing PS-like structural elements. We propose that genome packaging may be selectively initiated by the recognition of PSs comprising specific clustered structural motifs and synergistically promoted by non-specific RNA-protein interactions that cover the whole genome. However, it could be highly challenging to distinguish specific from non-specific RNA binding events for CoV N proteins. To study the RNA-N protein complex with resolved structural and spatial heterogeneity, microfluidic small-angle X-ray scattering could be a versatile technique for future studies to resolve the particle size, shapes, and dynamics of the RNA transcripts upon interacting with the N proteins. On the other hand, to pinpoint the locations and the structural features of the PS in SARS-CoV-2 gRNAs, further *in vivo* studies that take advantage of CoV reverse genetic systems to generate specific packaging-deficient mutants should be carried out.

Electrophoretic mobility shift assays have been widely used as a versatile tool to investigate RNA-protein interactions. However, this technique has inherent limitations, particularly in studying the dynamic and heterogeneous interactions between IDR-containing proteins and RNAs. These limitations can make it challenging to distinguish specific interactions without supplementary evidence. Despite these challenges, we emphasize that EMSA remains a crucial first-line tool for characterizing RNA-protein interactions, and have taken significant measures to interpret our EMSA and LLPS experiments appropriately. The insights gained from these assays laid the groundwork for more advanced structural approaches, such as cryo-electron microscopy (cryo-EM) and small-angle X-ray scattering (SAXS), which we propose as future steps to resolve the molecular details of RNA-N protein interactions. In conclusion, we have revealed previously uncharacterized RNA-protein binding features between SARS-CoV-2 5'-proximal gRNA transcripts and the N proteins, providing new perspectives on N protein-mediated genome packaging and

other regulatory functions. We highlight the critical roles played by SL5a/b/c, which influence the conformation of the 5'-proximal transcripts and affect the condensate formations with N-proteins. For the RNA-binding peptides identified in this study, further optimizations to improve their binding ability, structural selectivities, and physiological properties would be crucial for the development of prospective peptide antagonists against SARS-CoV-2.

Acknowledgments

We thank the Biomedical Resource Core at the First Core Labs, National Taiwan University College of Medicine, providing the cDNA amplicon of SARS-CoV-2. We thank Dr. S.-J. Liu, P. Chong, and their coworkers for providing the peptide library of N proteins. We also thank Dr. R.C.L. Olsthoorn for his pivotal contribution to initiate this work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the programmes of Research Center for Epidemic Prevention Science (MOST111-2321-B-006-009) and Columbus Young Scholar Fellowship (MOST109-2636-B-006-005) from Ministry of Science and Technology, Taiwan. National Cheng Kung University, Taiwan, and Higher Education Sprout Project, Ministry of Education to the Headquarters of University, Taiwan, also provide funding.

Author contributions

S.-C.C.: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review & Editing, Supervision, Project administration, Funding acquisition. C.-T.X.: Validation, Investigation, Formal analysis. C.-F.C.: Formal analysis, Investigation. C.-S.Y.: Investigation, P.-H.L.: Investigation. W.-M.L.: Methodology, Resources, Writing – Original Draft, Supervision. Y.C.: Resources, Supervision, Funding acquisition. C.-H.Y.: Conceptualization, Supervision, Writing – Review & Editing, Funding acquisition.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article or its supplementary materials.

ORCID

Shih-Cheng Chen  <http://orcid.org/0000-0002-4764-0510>
 Wei-Min Liu  <http://orcid.org/0000-0003-3607-1911>
 Yeh Chen  <http://orcid.org/0000-0002-7740-0446>
 Chien-Hung Yu  <http://orcid.org/0000-0003-0337-4473>

References

- [1] He R, Dobie F, Ballantine M, et al. Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem Biophys Res Commun.* 2004;316(2):476–483. doi: 10.1016/j.bbrc.2004.02.074
- [2] Yu IM, Gustafson CL, Diao J, et al. Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its c-terminal domain. *J Biol Chem.* 2005;280(24):23280–23286. doi: 10.1074/jbc.M501015200

- [3] Chen CY, Chang CK, Chang YW, et al. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol.* 2007;368(4):1075–1086. doi: [10.1016/j.jmb.2007.02.069](#)
- [4] Chang CK, Hsu YL, Chang YH, et al. Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol.* 2009;83(5):2255–2264. doi: [10.1128/JVI.02001-08](#)
- [5] Chang CK, Hou MH, Chang CF, et al. The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res.* 2014;103:39–50. doi: [10.1016/j.antiviral.2013.12.009](#)
- [6] Zhu N, Zhang D, Wang W, et al. China novel coronavirus, T. Research, a novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727–733. doi: [10.1056/NEJMoa2001017](#)
- [7] Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579(7798):270–273. doi: [10.1038/s41586-020-2012-7](#)
- [8] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* 2020;395(10224):565–574. doi: [10.1016/S0140-6736\(20\)30251-8](#)
- [9] Masters PS. Coronavirus genomic RNA packaging. *Virology.* 2019;537:198–207. doi: [10.1016/j.virol.2019.08.031](#)
- [10] Molenkamp R, Spaan WJM. Identification of a specific interaction between the coronavirus mouse hepatitis virus A59 nucleocapsid protein and packaging signal. *Virology.* 1997;239(1):78–86. doi: [10.1006/viro.1997.8867](#)
- [11] Kuo L, Masters PS. Functional analysis of the murine coronavirus genomic RNA packaging signal. *J Virol.* 2013;87(9):5182–5192. doi: [10.1128/JVI.00100-13](#)
- [12] Hsin WC, Chang CH, Chang CY, et al. Nucleocapsid protein-dependent assembly of the RNA packaging signal of middle east respiratory syndrome coronavirus. *J Biomed Sci.* 2018;25(1):47. doi: [10.1186/s12929-018-0449-x](#)
- [13] Chang CK, Sue SC, Yu TH, et al. The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure. *FEBS Lett.* 2005;579(25):5663–5668. doi: [10.1016/j.febslet.2005.09.038](#)
- [14] Huang Q, Yu L, Petros AM, et al. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry.* 2004;43(20):6059–6063. doi: [10.1021/bi036155b](#)
- [15] Ye Q, West AMV, Silletti S, et al. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* 2020;29(9):1890–1901. bioRxiv. doi: [10.1002/pro.3909](#)
- [16] Peng Y, Du N, Lei Y, et al. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *Embo J.* 2020;39(20):e105938. doi: [10.15252/embj.2020105938](#)
- [17] Khan A, Tahir Khan M, Saleem S, et al. Structural insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein. *Comput Struct Biotechnol J.* 2020;18:2174–2184. doi: [10.1016/j.csbj.2020.08.006](#)
- [18] Kang S, Yang M, Hong Z, et al. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B.* 2020;10(7):1228–1238. doi: [10.1016/j.apsb.2020.04.009](#)
- [19] Jayaram H, Fan H, Bowman BR, et al. X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J Virol.* 2006;80(13):6612–6620. doi: [10.1128/JVI.00157-06](#)
- [20] Zhang B, Tian J, Zhang Q, et al. Comparing the nucleocapsid proteins of human coronaviruses: structure, immunoregulation, vaccine, and targeted drug. *Front Mol Biosci.* 2022;9:761173. doi: [10.3389/fmolb.2022.761173](#)
- [21] Hsu JN, Chen JS, Lin SM, et al. Targeting the N-Terminus domain of the coronavirus nucleocapsid protein induces abnormal oligomerization via allosteric modulation. *Front Mol Biosci.* 2022;9:871499. doi: [10.3389/fmolb.2022.871499](#)
- [22] Schiavina M, Pontoriero L, Tagliaferro G, et al. The role of disordered regions in orchestrating the properties of multidomain proteins: the SARS-CoV-2 nucleocapsid protein and its interaction with enoxaparin. *Biomolecules.* 2022;12(9):1302. doi: [10.3390/biom12091302](#)
- [23] Pontoriero L, Schiavina M, Korn SM, et al. NMR reveals specific tracts within the intrinsically disordered regions of the SARS-CoV-2 nucleocapsid protein involved in RNA encountering. *Biomolecules.* 2022;12(7):929. doi: [10.3390/biom12070929](#)
- [24] Schiavina M, Pontoriero L, Uversky VN, et al. The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1–248 residue construct: sequence-specific resonance assignments through NMR. *Biomol NMR Assign.* 2021;15(1):219–227. doi: [10.1007/s12104-021-10009-8](#)
- [25] Dinesh DC, Chalupska D, Silhan J, et al. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* 2020;16(12):e1009100. doi: [10.1371/journal.ppat.1009100](#)
- [26] Hsieh PK, Chang SC, Huang CC, et al. Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J Virol.* 2005;79(22):13848–13855. doi: [10.1128/JVI.79.22.13848-13855.2005](#)
- [27] Luo H, Chen J, Chen K, et al. Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry.* 2006;45(39):11827–11835. doi: [10.1021/bi0609319](#)
- [28] Chen SC, Olsthoorn RCL. Group-specific structural features of the 5'-proximal sequences of coronavirus genomic RNAs. *Virology.* 2010;401(1):29–41. doi: [10.1016/j.virol.2010.02.007](#)
- [29] Chen SC, van den Born E, van den Worm SHE, et al. New structure model for the packaging signal in the genome of group IIa coronaviruses. *J Virol.* 2007;81(12):6771–6774. doi: [10.1128/JVI.02231-06](#)
- [30] Xing L, Zhang L, Kessel JV, et al. Identification of cis-acting sequences required for selective packaging of bovine adenovirus type 3 DNA. *J Gener Virol.* 2003;84(11):2947–2956. doi: [10.1099/vir.0.19418-0](#)
- [31] Terasaki K, Narayanan K, Makino S, et al. Identification of a 1.4-kb-long sequence located in the nsp12 and nsp13 coding regions of SARS-CoV-2 genomic RNA that mediates efficient viral RNA packaging. *J Virol.* 2023;97(7):e0065923. doi: [10.1128/jvi.00659-23](#)
- [32] Chen SC, Olsthoorn RCL, Yu CH. Structural phylogenetic analysis reveals lineage-specific RNA repetitive structural motifs in all coronaviruses and associated variations in SARS-CoV-2. *Virus Evol.* 2021;7(1):veab021. doi: [10.1093/ve/veab021](#)
- [33] Chaturvedi S, Vasen G, Pablo M, et al. Identification of a therapeutic interfering particle—a single-dose SARS-CoV-2 antiviral intervention with a high barrier to resistance. *Cell.* 2021;184(25):6022–6036.e18. doi: [10.1016/j.cell.2021.11.004](#)
- [34] Terasaki K, Makino S, Heise MT. Requirement of the N-terminal region of nonstructural protein 1 in cis for SARS-CoV-2 defective RNA replication. *J Virol.* 2024;98(9):e0090024. doi: [10.1128/jvi.00900-24](#)
- [35] Morales L, Mateos-Gomez PA, Capiscol C, et al. Transmissible gastroenteritis coronavirus genome packaging signal is located at the 5' end of the genome and promotes viral RNA incorporation into virions in a replication-independent process. *J Virol.* 2013;87(21):11579–11590. doi: [10.1128/JVI.01836-13](#)
- [36] Ko TP, Wang YC, Yang CS, et al. Crystal structure and functional implication of bacterial STING. *Nat Commun.* 2022;13(1):26. doi: [10.1038/s41467-021-26583-3](#)
- [37] Ko TP, Wang YC, Tsai CL, et al. Crystal structure and functional implication of a bacterial cyclic AMP–AMP–GMP synthetase. *Nucleic Acids Res.* 2021;49(8):4725–4737. doi: [10.1093/nar/gkab165](#)

- [38] Lan TCT, Allan MF, Malsick LE, et al. Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat Commun.* 2022;13(1):1128. doi: [10.1038/s41467-022-28603-2](https://doi.org/10.1038/s41467-022-28603-2)
- [39] Garcia-Moran E, Hernandez M, Abad D, et al. Putative secondary structure at 5'UTR as a potential antiviral target against SARS-CoV-2. *Rev Esp Quimioter.* 2022;35(2):204–212. doi: [10.37201/req/153.2021](https://doi.org/10.37201/req/153.2021)
- [40] Miao Z, Tidu A, Eriani G, et al. Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biol.* 2021;18(4):447–456. doi: [10.1080/15476286.2020.1814556](https://doi.org/10.1080/15476286.2020.1814556)
- [41] Wacker A, Weigand JE, Akabayov SR, et al. Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* 2020;48(22):12415–12435. doi: [10.1093/nar/gkaa1013](https://doi.org/10.1093/nar/gkaa1013)
- [42] Manfredonia I, Nithin C, Ponce-Salvatierra A, et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* 2020;48(22):12436–12452. doi: [10.1093/nar/gkaa1053](https://doi.org/10.1093/nar/gkaa1053)
- [43] Rangan R, Zheludev IN, Hagey RJ, et al. RNA genome conservation and secondary structure in SARS-CoV-2 and sars-related viruses: a first look. *RNA.* 2020;26(8):937–959. doi: [10.1261/rna.076141.120](https://doi.org/10.1261/rna.076141.120)
- [44] Sun L, Li P, Ju X, et al. In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell.* 2021;184(7):1865–1883.e20. doi: [10.1016/j.cell.2021.02.008](https://doi.org/10.1016/j.cell.2021.02.008)
- [45] Misra VK, Draper DE. On the role of magnesium ions in RNA stability. *Biopolymers.* 1998;48(2–3):113–135. doi: [10.1002/\(SICI\)1097-0282\(1998\)48:2<113::AID-BIP3>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0282(1998)48:2<113::AID-BIP3>3.0.CO;2-Y)
- [46] Luo H, Chen Q, Chen J, et al. The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *FEBS Lett.* 2005;579(12):2623–2628. doi: [10.1016/j.febslet.2005.03.080](https://doi.org/10.1016/j.febslet.2005.03.080)
- [47] Guan BJ, Wu HY, Brian DA. An optimal cis -replication stem-loop IV in the 5' untranslated region of the mouse coronavirus genome extends 16 nucleotides into open reading frame 1. *J Virol.* 2011;85(11):5593–5605. doi: [10.1128/JVI.00263-11](https://doi.org/10.1128/JVI.00263-11)
- [48] Raman S, Bouma P, Williams GD, et al. Stem-Loop III in the 5' untranslated region is a cis -acting element in bovine coronavirus defective interfering RNA replication. *J Virol.* 2003;77(12):6720–6730. doi: [10.1128/JVI.77.12.6720-6730.2003](https://doi.org/10.1128/JVI.77.12.6720-6730.2003)
- [49] Raman S, Brian DA. Stem-loop IV in the 5' untranslated region is a cis -acting element in bovine coronavirus defective interfering RNA replication. *J Virol.* 2005;79(19):12434–12446. doi: [10.1128/JVI.79.19.12434-12446.2005](https://doi.org/10.1128/JVI.79.19.12434-12446.2005)
- [50] Baric RS, Nelson GW, Fleming JO, et al. Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. *J Virol.* 1988;62(11):4280–4287. doi: [10.1128/jvi.62.11.4280-4287.1988](https://doi.org/10.1128/jvi.62.11.4280-4287.1988)
- [51] Nelson GW, Stohman SA, Tahara SM. High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *Microbiology.* 2000;81(1):181–188. doi: [10.1099/0022-1317-81-1-181](https://doi.org/10.1099/0022-1317-81-1-181)
- [52] Wu CH, Yeh SH, Tsay YG, et al. Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. *J Biol Chem.* 2009;284(8):5229–5239. doi: [10.1074/jbc.M805747200](https://doi.org/10.1074/jbc.M805747200)
- [53] Grosseohme NE, Li L, Keane SC, et al. Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts trs-cTRS RNA duplexes. *J Mol Biol.* 2009;394(3):544–557. doi: [10.1016/j.jmb.2009.09.040](https://doi.org/10.1016/j.jmb.2009.09.040)
- [54] Yang D, Leibowitz JL. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* 2015;206:120–133. doi: [10.1016/j.virusres.2015.02.025](https://doi.org/10.1016/j.virusres.2015.02.025)
- [55] Tan YW, Fang S, Fan H, et al. Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells. *Nucleic Acids Res.* 2006;34(17):4816–4825. doi: [10.1093/nar/gkl650](https://doi.org/10.1093/nar/gkl650)
- [56] Yang M, He S, Chen X, et al. Structural insight into the SARS-CoV-2 nucleocapsid protein C-Terminal domain reveals a novel recognition mechanism for viral transcriptional regulatory sequences. *Front Chem.* 2020;8:624765. doi: [10.3389/fchem.2020.624765](https://doi.org/10.3389/fchem.2020.624765)
- [57] Kuo L, Koetzner CA, Hurst KR, et al. Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid protein. *J Virol.* 2014;88(8):4451–4465. doi: [10.1128/JVI.03866-13](https://doi.org/10.1128/JVI.03866-13)
- [58] Ma Y, Tong X, Xu X, et al. Structures of the N- and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA-protein binding mechanism in coronavirus. *Protein Cell.* 2010;1(7):688–697. doi: [10.1007/s13238-010-0079-x](https://doi.org/10.1007/s13238-010-0079-x)
- [59] Zuniga S, Sola I, Moreno JL, et al. Coronavirus nucleocapsid protein is an RNA chaperone. *Virology.* 2007;357(2):215–227. doi: [10.1016/j.virol.2006.07.046](https://doi.org/10.1016/j.virol.2006.07.046)
- [60] Rafael Ciges-Tomas J, Franco ML, Vilar M. Identification of a guanine-specific pocket in the protein N of SARS-CoV-2. *Commun Biol.* 2022;5(1):711. doi: [10.1038/s42003-022-03647-8](https://doi.org/10.1038/s42003-022-03647-8)
- [61] Wang S, Dai T, Qin Z, et al. Targeting liquid-liquid phase separation of SARS-CoV-2 nucleocapsid protein promotes innate antiviral immunity by elevating MAVS activity. *Nat Cell Biol.* 2021;23(7):718–732. doi: [10.1038/s41556-021-00710-0](https://doi.org/10.1038/s41556-021-00710-0)
- [62] Perdikari TM, Murthy AC, Ryan VH, et al. SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins. *bioRxiv.* 2020;39(24): e106478. doi:[10.15252/embj.2020106478](https://doi.org/10.15252/embj.2020106478)
- [63] Iserman C, Roden CA, Boerneke MA, et al. Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol Cell.* 2020;80(6):1078–1091.e6. doi: [10.1016/j.molcel.2020.11.041](https://doi.org/10.1016/j.molcel.2020.11.041)
- [64] Wu W, Cheng Y, Zhou H, et al. The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. *Virol J.* 2023;20(1):6. doi: [10.1186/s12985-023-01968-6](https://doi.org/10.1186/s12985-023-01968-6)
- [65] Jack A, Ferro LS, Trnka MJ, et al. SARS-CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *PLoS Biol.* 2021;19(10):e3001425. doi: [10.1371/journal.pbio.3001425](https://doi.org/10.1371/journal.pbio.3001425)
- [66] Gruber AR, Lorenz R, Bernhart SH, et al. The vienna RNA websuite. *Nucleic Acids Res.* 2008;36(Web Server):W70–4. doi: [10.1093/nar/gkn188](https://doi.org/10.1093/nar/gkn188)
- [67] Rahman MS, Islam MR, Alam A, et al. Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *J Med Virol.* 2021;93(4):2177–2195. doi: [10.1002/jmv.26626](https://doi.org/10.1002/jmv.26626)
- [68] Matsuo T. Viewing SARS-CoV-2 nucleocapsid protein in terms of molecular flexibility. *Biology (Basel).* 2021;10(6):454. doi: [10.3390/biology10060454](https://doi.org/10.3390/biology10060454)
- [69] McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses.* 2014;6(8):2991–3018. doi: [10.3390/v6082991](https://doi.org/10.3390/v6082991)
- [70] Sarkar S, Runge B, Russell RW, et al. Atomic-resolution structure of SARS-CoV-2 nucleocapsid protein N-terminal domain. *J Am Chem Soc.* 2022;144(23):10543–10555. doi: [10.1021/jacs.2c03320](https://doi.org/10.1021/jacs.2c03320)
- [71] Korn SM, Lambert R, Furtig B, et al. 1H, 13C, and 15N backbone chemical shift assignments of the C-terminal dimerization domain of SARS-CoV-2 nucleocapsid protein. *Biomol NMR Assign.* 2021;15(1):129–135. doi: [10.1007/s12104-020-09995-y](https://doi.org/10.1007/s12104-020-09995-y)
- [72] de Luna Marques A, Caruso IP, Santana-Silva MC, et al. 1H, 15N and 13C resonance assignments of the N-terminal domain of the nucleocapsid protein from the endemic human coronavirus HKU1. *Biomol NMR Assign.* 2021;15(1):153–157. doi: [10.1007/s12104-020-09998-9](https://doi.org/10.1007/s12104-020-09998-9)
- [73] Ahamad S, Gupta D, Kumar V. Targeting SARS-CoV-2 nucleocapsid oligomerization: insights from molecular docking and molecular dynamics simulations. *J Biomol Struct Dyn.* 2022;40(6):2430–2443. doi: [10.1080/07391102.2020.1839563](https://doi.org/10.1080/07391102.2020.1839563)
- [74] Lo YS, Lin SY, Wang SM, et al. Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. *FEBS Lett.* 2013;587(2):120–127. doi: [10.1016/j.febslet.2012.11.016](https://doi.org/10.1016/j.febslet.2012.11.016)

- [75] Zúñiga S, Cruz JLG, Sola I, et al. Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J Virol*. 2010;84(4):2169–2175. doi: [10.1128/JVI.02011-09](https://doi.org/10.1128/JVI.02011-09)
- [76] Zachrdla M, Savastano A, Ibanez de Opakua A, et al. Contributions of the N-terminal intrinsically disordered region of the severe acute respiratory syndrome coronavirus 2 nucleocapsid protein to RNA-induced phase separation. *Protein Sci*. 2022;31(9):e4409. doi: [10.1002/pro.4409](https://doi.org/10.1002/pro.4409)
- [77] Wu C, Qavi AJ, Hachim A, et al. Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain. *iScience*. 2021;24(6):102681. doi: [10.1016/j.isci.2021.102681](https://doi.org/10.1016/j.isci.2021.102681)
- [78] Azad GK. Identification and molecular characterization of mutations in nucleocapsid phosphoprotein of SARS-CoV-2. *PeerJ*. 2021;9:e10666. doi: [10.7717/peerj.10666](https://doi.org/10.7717/peerj.10666)
- [79] Zhao H, Nguyen A, Wu D, et al. Plasticity in structure and assembly of SARS-CoV-2 nucleocapsid protein. *bioRxiv*. 2022;1(2). doi:[10.1093/pnasnexus/pgac049](https://doi.org/10.1093/pnasnexus/pgac049)
- [80] Rozycki B, Boura E. Conformational ensemble of the full-length SARS-CoV-2 nucleocapsid (N) protein based on molecular simulations and SAXS data. *Biophys Chem*. 2022;288:106843. doi: [10.1016/j.bpc.2022.106843](https://doi.org/10.1016/j.bpc.2022.106843)
- [81] Ribeiro-Filho HV, Jara GE, Batista FAH, et al. Structural dynamics of SARS-CoV-2 nucleocapsid protein induced by RNA binding. *PLoS Comput Biol*. 2022;18(5):e1010121. doi: [10.1371/journal.pcbi.1010121](https://doi.org/10.1371/journal.pcbi.1010121)
- [82] Redzic JS, Lee E, Born A, et al. The inherent dynamics and interaction sites of the SARS-CoV-2 nucleocapsid N-Terminal region. *J Mol Biol*. 2021;433(15):167108. doi: [10.1016/j.jmb.2021.167108](https://doi.org/10.1016/j.jmb.2021.167108)
- [83] Cai T, Yu Z, Wang Z, et al. Arginine methylation of SARS-Cov-2 nucleocapsid protein regulates RNA binding, its ability to suppress stress granule formation, and viral replication. *J Biol Chem*. 2021;297(1):100821. doi: [10.1016/j.jbc.2021.100821](https://doi.org/10.1016/j.jbc.2021.100821)
- [84] Wu CH, Chen PJ, Yeh SH. Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host & Microbe*. 2014;16(4):462–472. doi: [10.1016/j.chom.2014.09.009](https://doi.org/10.1016/j.chom.2014.09.009)
- [85] Keane SC, Liu P, Leibowitz JL, et al. Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucleocapsid (N) protein. *J Biol Chem*. 2012;287(10):7063–7073. doi: [10.1074/jbc.M111.287763](https://doi.org/10.1074/jbc.M111.287763)
- [86] Cubuk J, Alston JJ, Incicco JJ, et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun*. 2021;12(1):1936. doi: [10.1038/s41467-021-21953-3](https://doi.org/10.1038/s41467-021-21953-3)
- [87] Syed AM, Taha TY, Tabata T, et al. Rapid assessment of SARS-CoV-2-evolved variants using virus-like particles. *Science*. 2021;374(6575):1626–1632. doi: [10.1126/science.abc6184](https://doi.org/10.1126/science.abc6184)
- [88] Roden CA, Dai Y, Giannetti CA, et al. Double-stranded RNA drives SARS-CoV-2 nucleocapsid protein to undergo phase separation at specific temperatures. *Nucleic Acids Res*. 2022;50(14):8168–8192. doi: [10.1093/nar/gkac596](https://doi.org/10.1093/nar/gkac596)
- [89] Mohammadi-Dehcheshmeh M, Moghbeli SM, Rahimirad S, et al. A transcription regulatory sequence in the 5' untranslated region of SARS-CoV-2 is vital for virus replication with an altered evolutionary pattern against human inhibitory MicroRNAs. *Cells*. 2021;10(2):319. doi: [10.3390/cells10020319](https://doi.org/10.3390/cells10020319)
- [90] Raden M, Ali SM, Alkhnbashi OS, et al. Freiburg RNA tools: a central online resource for RNA-focused research and teaching. *Nucleic Acids Res*. 2018;46(W1):W25–W29. doi: [10.1093/nar/gky329](https://doi.org/10.1093/nar/gky329)
- [91] Smith C, Heyne S, Richter AS, et al. Freiburg RNA tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res*. 2010;38(Web Server):W373–7. doi: [10.1093/nar/gkq316](https://doi.org/10.1093/nar/gkq316)

Appendix

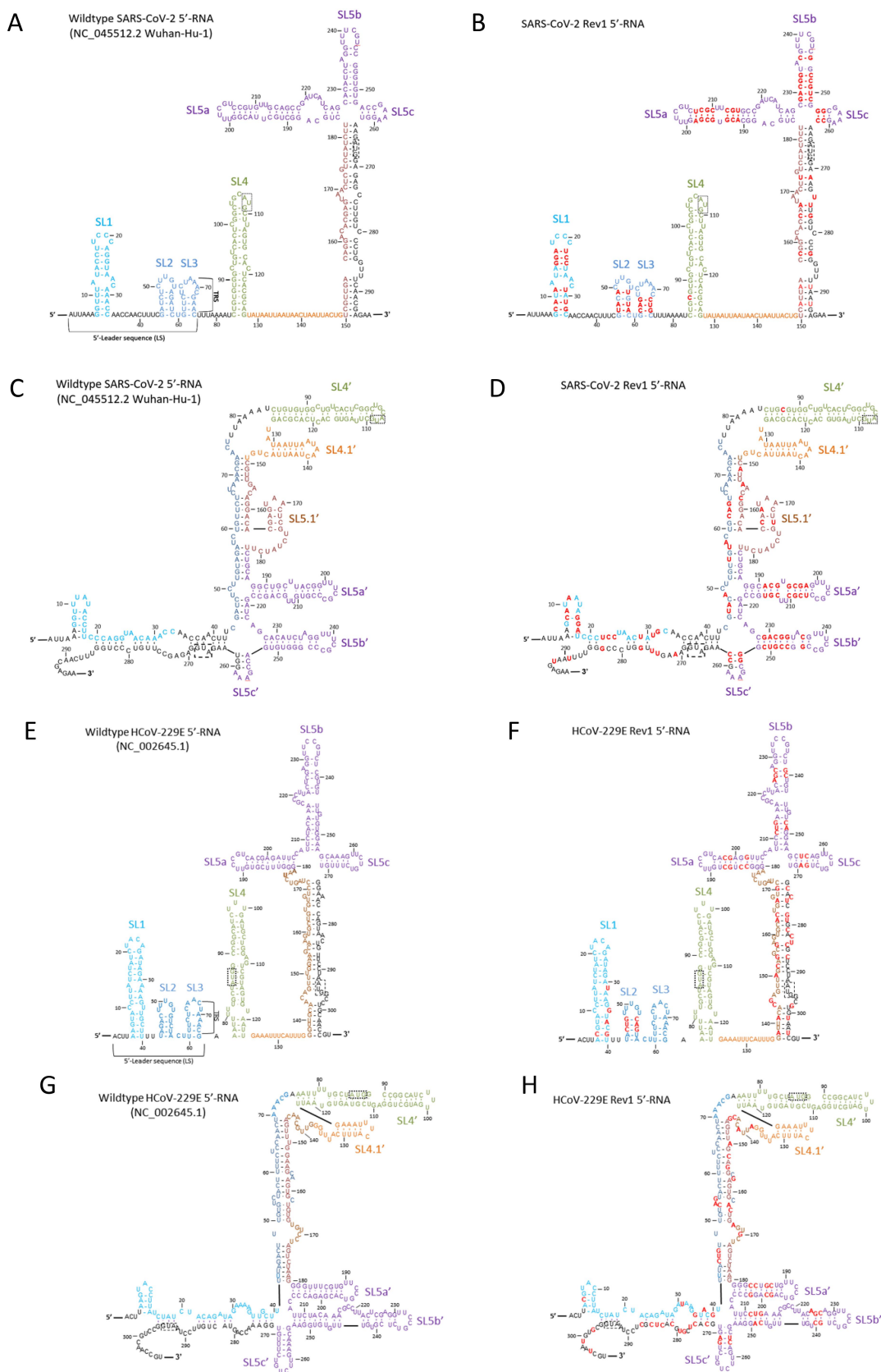


Figure A1. The canonical and the potential alternative secondary structure models of SARS-CoV-2 5'-proximal gRNA sequence. The canonical structure models of (A) Wildtype and (B) the mutant Rev 1 are shown, in comparison with their potential folding of alternative conformations (C and D), respectively. The alternative conformations are adapted from model proposed by [89].

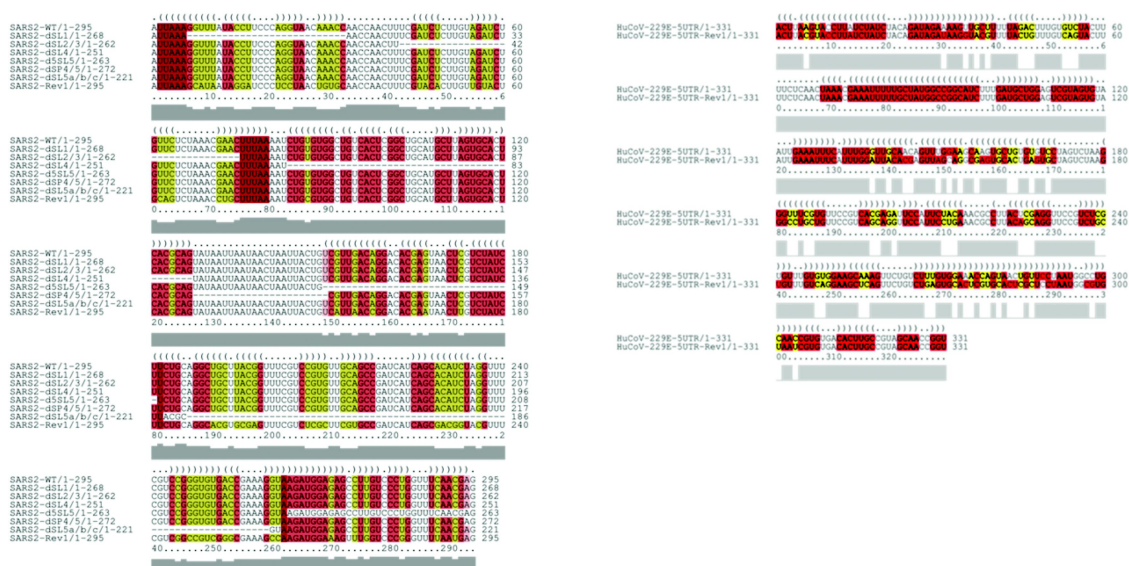


Figure A2. Sequences of CoV 5'-transcripts studied in this report. The sequences of wildtype and mutant 5'-transcripts were aligned and predicted for their consensus secondary structures with LocARNA provided by Freiburg RNA Tools. Red color-coded sequences represent highly conserved pairing nucleotides, while yellow-coded were less conserved in LocARNA's default algorithm. [88, 90, 91].

Table A1. Lists of the peptide sequences of the peptide library. Eighty 15-mer peptides with 10-a.a. overlaps spanning the whole 646 SARS-CoV N protein were synthesized. The charges of each a.a. (at pH7.5) were indicated positive (+), neutral (0), and negative (-), 647 respectively. Net charges were accumulatively calculated for each peptide as indicated. Significant RNA-interacting peptides were 648 labeled as "●" while peptides with no significant RNA binding activities were labeled as "○" according to the results showed 649 in Figure 6.

No.	Sequence	Start position	Charge distribution	Net charge	RNA binding	No.	Sequence	Start position	Charge distribution	Net charge	RNA binding
1	MSDNGPQSNQRSAPR	1	00-0000000+000+	1	○	41	GSSRGNSPARMASGG	201	000+00000+00000	2	○
2	PQSNQRSAPRITFGG	6	00000+000+00000	2	○	42	NSPARMASGGGETAL	206	0000+000000-000	0	○
3	RSAPRITFGGPTDST	11	+000+0000000-00	1	○	43	MASGGGETALALLLL	211	000000-00000000	-1	○
4	ITFGGPTDSTDNNGN	16	0000000-00-0000	-2	○	44	GETALALLLLDRLNQ	216	0-00000000+000	-1	○
5	PTDSTDNNGNGGRNG	21	00-00-000000+00	-1	○	45	ALLLLDRLNLQLESKV	221	00000+0000-0+0	0	○
6	DNNGNGGRNGARPKQ	26	-000000+000+0+0	2	○	46	DRLNLQLESKVSGKGQ	226	+0000-0+000+00	1	○
7	GGRNGARPKQRRPQG	31	00+000+0+0+0000	5	●	47	LESKVSGKGQQQQGG	231	0-0+000+0000000	1	○
8	ARPKQRRPQGLPNNT	36	0+0+0+0+0000000	4	●	48	SGKGQQQQGGQTVTKK	236	00+0000000000++	3	○
9	RRPQGLPNNTASWFT	41	+0+000000000000	2	○	49	QQQGGQTVTKKSAEA	241	00000000+000-0	1	○
10	LPNNTASWFTALTOH	46	00000000000000	0	○	50	TVTKKSAEASAKPR	246	000+000-00+0+0	4	○
11	ASWFTALTOHKGKEEL	51	00000000000+0	-1	○	51	SAEASAKPRQKRTA	251	000-00+0+0+000	4	●
12	ALTOHKGKEELRFPFG	56	000000+0-0+0+0	1	○	52	SKPRQKRTATKQYN	256	0+0+0+0+000+000	6	○
13	GKEELRFPFGQGVPI	61	0+0-0+0+000000	1	○	53	QKRTATKQYNVTQAF	261	0+000+00000000	3	○
14	RFPFGQGVPIINTNSG	66	+00+00000000000	2	○	54	TKQYNVTQAFGRRGP	266	0+00000000+00	3	○
15	QGVPIINTNSGPDQI	71	00000000000-00	-2	○	55	VTQAFGRRGPEQTQGG	271	000000+00-0000	1	○
16	INTNSGPDQIQYYRR	76	000000-00000+0	0	○	56	GRRGPEQTQGNFGDQ	276	0+0+00-000000-0	0	○
17	PDDQIQYYRRATRRV	81	0-00000+00+00	2	●	57	EQTGQNFQDQLIRQ	281	-0000000-0-00+0	-2	○
18	GYRRATRRVRGDDG	86	000+00+0+0+00-0	4	○	58	NFGDQLIRQGTQDYK	286	000-0-00+000-0+	1	○
19	ATRRVRGDDGKMKEEL	91	00+0+00-0+0+0	3	○	59	DLIRQGTQDYKHWPQI	291	-00+000-0+00000	0	○
20	RGGDDGKMKEELSPRWY	96	+00-0+0+000+00	2	○	60	GTQDYKHWPQIAQFAP	296	00-0+0000000000	0	○
21	KMKELSPRWYFYFLG	101	+0+0-000+000000	2	○	61	HWPQIAQFAPSASAF	301	00000000000000	0	○
22	SPRWYFYFLGTGPEA	106	00+00000000000	0	○	62	AQFAPSASAFFGMSR	306	00000000000000+	1	○
23	FYFLGTGPEASLPYG	111	00000000-000000	-1	○	63	SASAFFGMSRIGMEV	311	000000000+000-0	0	○
24	TGPEASLPYGANKKEG	116	000-00000000+0	-1	○	64	FGMSRIGMEVTPSGT	316	0000+000-000000	0	○
25	SLPYGANKEGIVVWA	121	0000000+000000	0	○	65	IGMEVTPSGTWLTYH	321	000-00000000000	-1	○
26	ANKKEGIVVWATEGAL	126	00+0000000-000	-1	○	66	TPSGTWLTYHGAIKL	326	0000000000000+0	1	○
27	IVVWATEGALNTPKD	131	000000-000000+-	-1	○	67	WLTYHGAIKLDDKDP	331	00000000+0-+0	-1	○
28	TEGALNTPKDHIGTR	136	0-00000+0000+0	0	○	68	GAIKLDDKDPQFQDN	336	000+0-+000+0+	-1	○
29	NTPKDHIGTRNPNNN	141	000+0000+00000	1	○	69	DDKDPQFQDNVILLN	341	-+000+000000	-2	○
30	HIGTRNPNNNATVLL	146	0000+000000000	1	○	70	QFQDNVILLNKHIDA	346	00+000000+00-0	0	○
31	NPNNNATVLLQLPQG	151	00000000000000	0	○	71	VILLNKHIDAYKTFP	351	00000+00-00+000	1	○
32	AATVLLQLPQGTTLPK	156	0000000000000+	1	○	72	KHIDAYKTFPPTPEK	356	+00-00+00000-0+	1	○
33	QLPQGTTLPKGFYAE	161	000000000+0000	0	○	73	YKTFPPTPEKPKDKKK	361	0+00000-0++000	4	○
34	TLPKGFYAEGRSGG	166	0000+0000-00+00	1	○	74	PTEPKDKKKKTDEA	366	00-0++000+0-0	2	○
35	GFYAEGRSGGQSSAS	171	0000-00+000000	0	○	75	KDKKKKTDEAQLPQ	371	+0000-0000000	2	○
36	GRSGGQSSASRSSRS	176	00+0000000+000+	3	○	76	KTDEAQLPQQRQKKQ	376	+0-000000+0+00	2	○
37	SQASRSSRSRSGNS	181	00000+000+0+000	3	●	77	QLPQQRQKKQPTVTL	381	00000+0+000000	3	○
38	RSSRSRSGNSRNPST	186	+000+0+000+0000	4	○	78	RQKKQPTVTLPAAD	386	+0+00000000000	2	○
39	SRGNSRNPSTPGSSRG	191	0+000+0000000+0	3	○	79	PTVTLPAADMDDFS	391	000000000-0-00	-3	○
40	RNPSTPGSSRGNPAR	196	+0000000+00000+	3	○	80	LPAADMDDFSRLQNL	396	0000-0-00+0000	-2	○

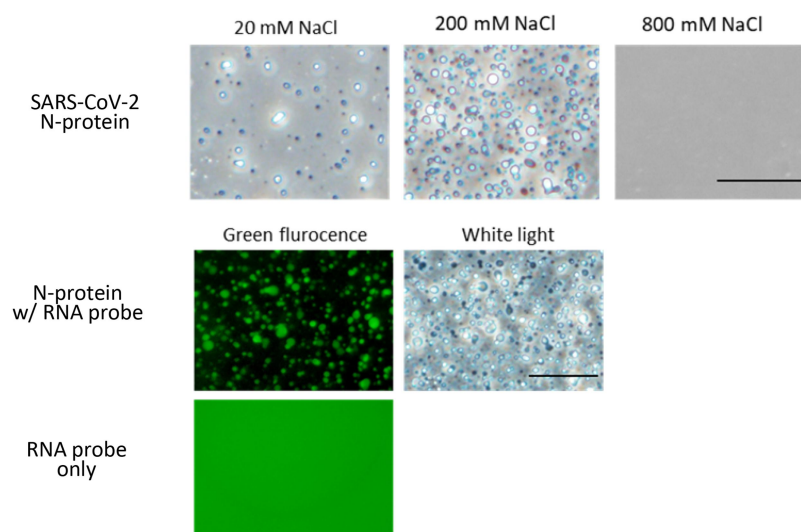


Figure A3. N-protein condensates *in vitro*. (A) Purified SARS-CoV-2 N-proteins were tested for condensate formation in different concentration of NaCl. The N-protein condensates were co-localized with the fluorescence-labeled RNA probes. Scale bar: 100 μ m.